Improvement in universal PIV outlier detection by means of coherence adaptivity

Alessandro Masullo[1], Raf Theunissen[1]

[1]Department of Aerospace Engineering, University of Bristol, Bristol, United Kingdom
a.masullo@bristol.ac.uk

INTRODUCTION

Particle Image Velocimetry (PIV) is a fairly well established technique for the flow measurements that finds applications in many fields. It allows the estimation of the average flow velocity in small sub-areas, between sequential images, by means of statistical operators. The most common operator in PIV is the cross-correlation, known for its versatility and robustness. Although widely used, its accuracy is strongly influenced by image quality, particle image density, underlying flow features, and velocity gradients. In circumstances as a strong distortion of the particle patterns between two snapshots, or an insufficient number of particles, the velocity estimates can be strongly biased or completely wrong. As explored by Westerweel [1], although velocity outliers can be reduced up to less than 5% through carefully designed experiments [2], their occurrence in the results is almost inevitable and, in the case they were absent, a higher spatial accuracy could have been reached. The presence of outliers in the velocity field also strongly precludes the possibility to evaluate derived quantities, especially when first or second order derivatives are required, and for this reason vector outlier detection has received considerable attention within the PIV community.
Keane and Adrian [2] define the detectability of the measurement within a correlation window as the ratio of the first tallest peak to the second tallest peak of the correlation map. Though this can be useful to identify potential outliers it is insufficient to detect spurious vectors in the final velocity field. Instead, obtained velocity fields must undergo a secondary validation process. Many are the post-interrogation algorithms that have been implemented in the last two decades. Due to the instantaneous nature of the flow measurements typically analysed with PIV, most of these techniques rely on a spatial comparison of each vector with its neighbours at the same time step. One of the simplest algorithm is the "manual" method [3], where a vector is considered to be wrong where its magnitude and direction are different from their closest neighbours according to a tolerance that is "manually" set. Westerweel [1] also studied the potential of a statistical approach to detect spurious vectors, investigating and comparing three different methods: the global-mean, the local-mean and the local-median, showing the latter having the highest efficiency. Song et al. [4] realised that although the median filter proposed by Westerweel is able to remove most of the outliers, correct vectors are affected by a smoothing effect due to false vectors. The authors therefore proposed an outlier detection based on the Delaunay Triangulation and continuity equation. Liang et al. [5] proposed a vector validation algorithm based on an artificial neural network and studied the influence of velocity gradients on the percentage of detections. An interesting approach based on human vision has been proposed by Reiz et al. [6] where an algorithm is presented based on the principles of Gestalt psychology and the recognition of continuous curves.
One of the most sensitive parts in the vector validation is the distinction between outliers and vectors that look different from their neighbours because of high gradients. To mitigate this problem, an universal outlier detection for PIV data has been proposed by Westerweel and Scarano [7], aiming to extend the median test [1] to more general turbulent flows by means of a normalized median test (NMT). More recently, Duncan et al. [8] highlighted that this approach is unsuitable for unstructured data, such as in Particle Tracking Velocimetry (PTV) or spatial adaptive routines [9], and suggest to include a distance weighting system.
In spite of the number of existing vector validation algorithms, a really important issue that has never been outlined and dealt with appropriately is the case of clusters of outliers. As stated before, erroneous velocity vectors often occur where images are noisy or do not contain enough information; in fact, should and area of the flow be low-seeded, it will probably cause outliers to occur in clusters. As far as the authors knowledge, all of the existing vector validation methodologies compare each vector with a constant and small number of neighbours and because of this limit in the "field of view", none of them will ever be able to detect big clusters of outliers. Furthermore, a small constant number of neighbours can be strongly insufficient to detect the validity of a vector, especially in case of strong gradients, causing some vectors to be over-detected. This problem is highlighted in Figure 1: on the left side of the figure there is a vector with its closest neighbours: looking at a small number of neighbours it is almost impossible to understand if it is correct, in fact, Figure 1a), shows how these vectors could be perfectly part of a coherent structure with strong gradients, while Figure 1b) depicts the case where the same vectors are just part of a bigger cluster of outliers.

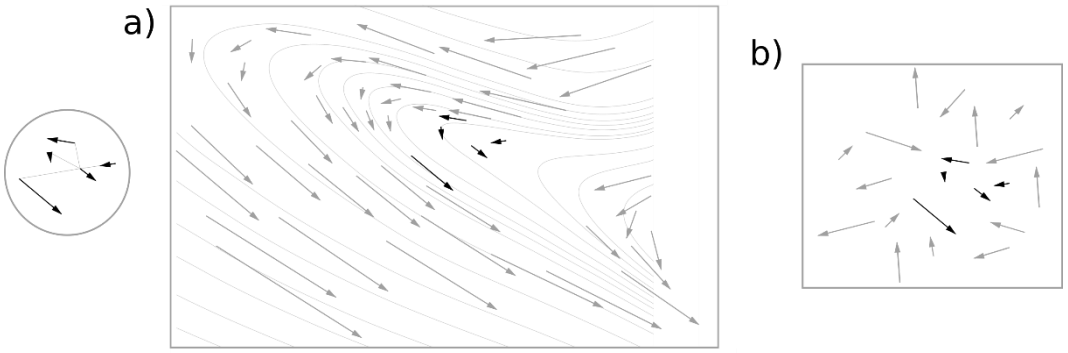


Figure 1 A small detail of some ambiguous neighbouring vectors. An example of these vectors as part of **a)** a coherent structure and **b)** a cluster of outliers

In this paper, the authors propose a novel algorithm for the vector validation in which the comparison of each vector is evaluated with a variable number of neighbours that is dynamically changed according to the coherence of the flow. This method not only allows to detect and remove big clusters of outliers from the velocity field, but it also allows to decrease the over-detection in case of strong gradients (i.e. correct vectors detected as outliers). To further improve the detection, the algorithm has been implemented with others enhancements; (1) dynamically variable number of neighbours, (2) a distance-based Gaussian weighting system, (4) a new approach for the median estimation and (5) magnitude/direction comparison. Furthermore, this method is applicable both on structured and unstructured grids, making it perfectly suitable for classic PIV/PTV algorithms or any kind of spatially adaptive algorithm. To assess the accuracy of this new algorithm in discerning correct vectors from outliers, a Monte Carlo simulation has been performed and results are shown in the next sections of this paper. An application to PIV has also been implemented to show the impact of this new vector validation on the analysis of experimental flow images.

METHODOLOGY

Human vision can be considered the best tool in discerning outliers from legitimate vectors and its strength probably relies in the ability to recognize coherent structures. Human detection starts with a comparison of a vector with its closest vicinity and continues extending the neighbourhood until the scrutinised vector is found to be part of a coherent region. Obviously, human detection becomes unaffordable when the number of velocity fields to be checked is in the order of thousands or more. For this reason, the aim of this new algorithm is to emulate human vision through a coherence-adaptive algorithm which recognizes consistent areas in the flow and compares each vector with a variable number of neighbours adapted to local flow field coherence.
The concept of coherence can be tricky to define, especially when dealing with scattered values. In a previous work [10], the coherence is defined as the sum of the differences of a vector and its neighbours, normalized by the sum of the modules of the neighbours. Although this definition of coherence allows to distinguish coherent areas where the velocity field is smooth enough, it is unable to discriminate a region of strong gradients from a cluster of false vectors, because it does not take into account vector positions.
In this paper, the authors argue the condition for a vector to be considered coherent, is that it fits a quadratic surface with its eight closest neighbours. It is important to underline that this coherence is not used to discern correct vectors from outliers. Instead it is only used to automatically predict the number of neighbours with which each vector must be compared. Once this number is known, all of the neighbours are used in the test, whether they are coherent or not. For sake of clarity, a group of non-coherent vectors, as those depicted in Figure 1, will be compared with a higher number of neighbours in order to understand if they are correct (Figure 1a) or they are false (Figure 1b). Therefore, the coherence function cannot be used to discern correct vectors from outliers but it allows to vary the number of neighbours in order to detect clusters of outliers.
The new validation algorithm proposed in this paper has been developed based on the median normalised threshold (NMT) test [7]: in a similar manner, a normalised residual is evaluated for each vector of the field based on the fluctuation of its neighbours but, in our case, the analysis is evaluated with a variable number of neighbours and the fluctuation is evaluated comparing magnitude and direction of each vector instead of the components.
The algorithmic sequence of the proposed validation routine are summarised as:

1. Coherence check: Evaluating the coherence for each vector of the field through a quadratic regression
2. Evaluate necessary number of neighbours: Estimating the minimum number of neighbours necessary for a reliable comparison, based on the coherence
3. Calculate the normalized residual: Evaluating the normalized residual of each vector cycling across the variable number of neighbours of the previous step
4. Validate vectors: Discriminating outliers and legitimate vectors by means of a universal threshold for the

normalized residual evaluated in the previous step

COHERENCE FUNCTION

As already stated in the introduction, a vector is considered coherent if it fits a quadratic surface with its eight closest neighbours. In particular, the coherence of a vector is defined as the residual of its components compared to the parabolic regression of its eight closest neighbours, conveniently normalized. Considering the vector $\mathbf{V}_0 = (u_0, v_0)$ located at $\mathbf{X}_0 = (x_0, y_0)$ and its neighbours $\mathbf{V_1}, \ldots, \mathbf{V}_8$ as depicted in Figure 2; the objective is to fit a parabolic curve for the components of these vectors using the least squares. For the horizontal components, it follows that the parabolic function $\Phi(x, y)$ is defined as:

$$\Phi_u(x, y) = a_0 + a_1 x + a_2 y + a_3 xy + a_4 x^2 + a_5 y^2 = u_i \tag{1}$$

where a_i are the unknown coefficients. Equation (1) can be rewritten for the components of every vector, yielding:

$$\begin{bmatrix} 1 & x_0 & y_0 & x_0 y_0 & {x_0}^2 & {y_0}^2 \\ 1 & x_1 & y_1 & x_1 y_1 & {x_1}^2 & {y_1}^2 \\ & & & \cdots & & \\ 1 & x_8 & y_8 & x_8 y_8 & {x_8}^2 & {y_8}^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ \vdots \\ a_6 \end{Bmatrix} = \begin{Bmatrix} u_0 \\ u_1 \\ \vdots \\ u_8 \end{Bmatrix} \tag{2}$$

or in matrix form;

$$X \cdot a = u \tag{3}$$

Since the parabolic fitting is performed with nine elements only, a single outlier in the neighbourhood may produce a completely wrong parabolic prediction, leading to an erroneous coherence value. To avoid this behaviour, vector components are weighted according to a Gaussian function of their variance:

$$w_i = exp\left[-\frac{{\sigma_i}^2}{2\left(\sum_i \frac{\sigma_i}{N} + \varepsilon \right)^2} \right] \tag{4}$$

$$\sigma_i = \sqrt{\left(u_i - \text{med}(u_j)\right)^2 + \left(v_i - \text{med}(v_j)\right)^2}, \; j = 0, ..., 8 \tag{5}$$

where ε is a background error (in PIV [7] this value is usually set to 0.1) while $\text{med}()$ is the median operator. Using these weights, the fitting equation becomes:

$$\left(X^T W X\right) a = \left(X^T W\right) u \tag{6}$$

where W is the matrix of the weights:

$$W_{ij} = 0 \text{ when } i \neq j, \; W_{ii} = w_i \tag{7}$$

Once the fitting has been performed and coefficient a_i evaluated, vector $\mathbf{V}_0$ can be compared with its parabolic fitting $\mathbf{\Phi}(x_0, y_0)$: the coherence C is the residual of the scrutinised vector's components normalized with the median of the magnitudes of the vectors:

$$C_x = \left[\frac{\Phi_u(x_0, y_0) - u_0}{\mathrm{med}\left(\sqrt{u_i^2 + v_i^2}\right) + \varepsilon} \right]^2 \tag{8}$$

$$C_y = \left[\frac{\Phi_v(x_0, y_0) - v_0}{\mathrm{med}\left(\sqrt{u_i^2 + v_i^2}\right) + \varepsilon} \right]^2 \tag{9}$$

The final coherence value is a simple mean of the horizontal and vertical coherence.

$$C = \frac{C_x + C_y}{2} \tag{10}$$

Once the coherence has been evaluated for all the vectors, a threshold allows to differentiate coherent vector from non-coherent ones. The threshold has been set to 0.1, so that vectors exceeding 10% of the residual are considered non-coherent and vice-versa. This threshold can be changed to adjust the sensitivity of the number of neighbours to the coherence of the flow, though the susceptibility of the algorithm to this parameter has been experimentally shown to be low.

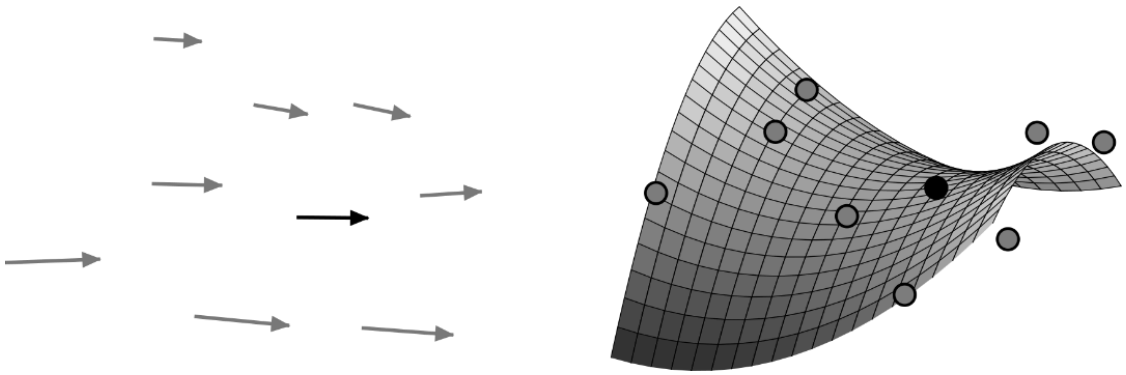


Figure 2 A vector with its eight closest neighbours and a parabolic fitting of the horizontal components

NUMBER OF NEIGHBOURS

The coherence adaptive variable number of neighbours of this algorithm is the key for detecting clusters of outliers. In fact, the size of the neighbourhood cannot be set just arbitrarily large, since regions where turbulence or strong gradients occur would be over-detected[1]; on the other hand, a small number of neighbours can be insufficient to understand if a vector differs from its neighbours because of strong gradients or because it is part of a cluster of false vectors. Consequently, the number of neighbours must be higher only where outliers are detected to be clustered, that is, in regions where less vectors are coherent.
Since the algorithm employs a modified version of median to discern outliers from correct vectors, the authors argue that at least 50% of the vectors in the neighbourhood of the scrutinised vector need to be coherent for a reliable comparison.

[1] An arbitrarily high number of neighbours would lead a vortex in a uniform flow to be considered as a "cluster of outliers"

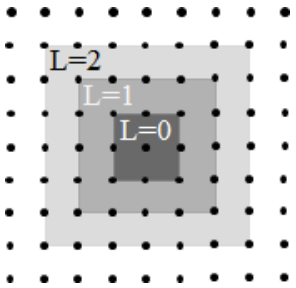


Figure 3 Level of neighbourhood on a structured grid

The variable number of neighbours with which each vector must be compared is adjusted through the level of neighbourhood L. A value of L = 1 is equivalent to the eight closest neighbours, L = 2 consists of the eight neighbours of the eight closest neighbours, and so on. On a structured grid, the number of neighbours N can be written as a function of the level of neighbours L as it follows:

$$N = (2L+1)^2 - 1 \tag{11}$$

To incorporate the dependency between vectors in case of correlation window overlap (WOR), a minimum value of L is imposed, which varies with WOR; L_{min}=1-ceil([1-WOR]$^{-1}$). Using the coherence evaluated in the previous step, the level of neighbourhood for each vector of the field is progressively increased until half of the neighbours of each vector are coherent. The result of this process is shown in Figure 4, where a cluster of outliers is surrounded by coherent vectors: the level of neighbourhood gradually increases going towards the centre of the cluster, so that each outlier can be compared with an adequate number of coherent neighbours, even though it is surrounded by outliers.

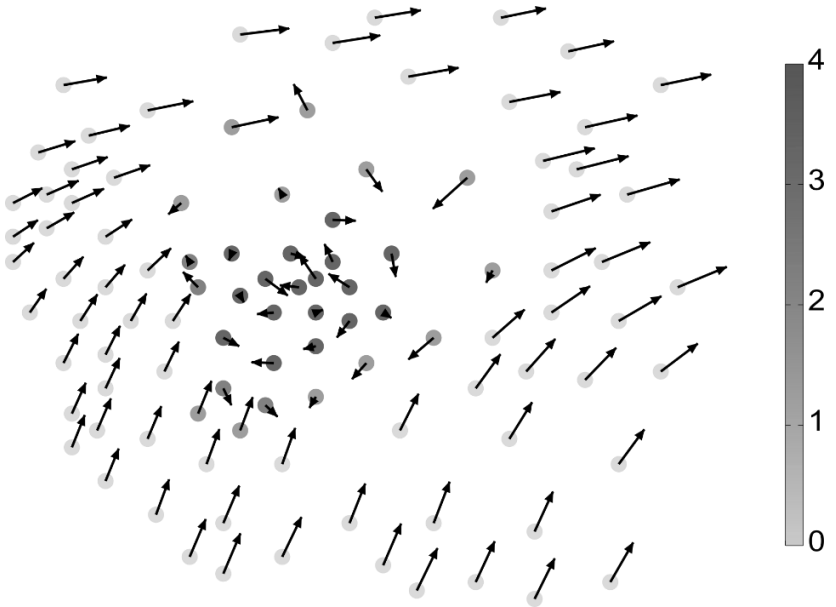


Figure 4 Level of neighbourhood evaluated in vicinity of a cluster of outliers. Notice how the level of neighbourhood increases going towards the centre of the cluster.

NORMALIZED RESIDUAL

The variable number of neighbours discussed in the previous section can be straightforwardly implemented in some of the already existing vector validation algorithms, like the Normalised Median Threshold (NMT [7]) or the Distance Weighted Normalised Median Threshold (DW-NMT [8]), and results for these algorithms are presented in the next sections.

In this section the authors propose an improved algorithm to evaluate the normalized residual, based on the NMT, which allows to further decrease the number of over-detections. In particular, this result is accomplished by means of three different innovations; (1) comparison of direction and magnitude instead of vector components, (2)a distance-based Gaussian weighting system and (3) a new approach for the median estimation

The normalized residual for a vector $\mathbf{V_0} = (u_0, v_0)$ is the combination of its magnitude and direction residuals:

$$R_0 = \sqrt{r_\alpha{}^2 + r_n{}^2} \tag{12}$$

$$r_{n,\alpha} = \frac{\mathrm{fluc}_0^{n,\alpha}}{\mathrm{fluc}_{med}^{n,\alpha} + \varepsilon^{n,\alpha}} \tag{13}$$

where r_n and r_α are the normalized fluctuation of magnitude and direction, fluc_0 is the fluctuation between $\mathbf{V_0}$ and its neighbours prediction, fluc_{med} is the average fluctuation of its neighbours, while ε is the background error, related to the measurement accuracy. Although the expressions of the normalized fluctuation for magnitude and angle have the same form, some differences must be pointed out. Vector $\mathbf{V_0}$ is surrounded by a certain number of neighbours N_0, namely $\mathbf{V_i} = (u_i, v_i)\ i = 1, 2, \ldots, N_0$. First, the median of $\mathbf{V_i}$ magnitudes is calculated as:

$$n_{med} = \sqrt{\mathrm{amed}(u_i, w_i)^2 + \mathrm{amed}(v_i, w_i)^2} \tag{14}$$

then the magnitude of $\mathbf{V_0}$:

$$n_0 = \sqrt{u_0^2 + v_0^2} \tag{15}$$

so that the magnitude fluctuation of $\mathbf{V_0}$ can be evaluated as:

$$\mathrm{fluc}_0^n = |n_0 - n_{med}| \tag{16}$$

The function *amed* is the averaged-weighted median, described afterwards. Each neighbouring vector $\mathbf{V_i}$ is weighted according to a Gaussian function of d_i, the distance of $\mathbf{V_i}$ from $\mathbf{V_0}$:

$$w_i = \exp\left[-\left(\frac{d_i}{\sigma}\right)^2\right] \tag{17}$$

where σ is set to 1.24 in accordance with Agüí and Jiménez [11]:

$$\sigma = 1.24 \cdot \frac{1}{N_0} \sum_{i=1}^{N_0} d_i \tag{18}$$

The fluctuation of the neighbours fluc_{med} is evaluated in a similar manner, but using the median instead of the averaged weighted median:

$$\mathrm{fluc}_{med}^n = \mathrm{med}(|n_i - n_{med}|) \tag{19}$$

since the median norm n_{med} is already evaluated by means of the weighted median, the fluctuation doesn't need to be weighted twice. The normalized fluctuation of the angle is evaluated in a similar manner, but the angle fluctuation and the median angle must be defined first. The angle fluctuation of $\mathbf{V_0}$ is defined as:

$$\mathrm{fluc}_0^\alpha = \min\{2\pi - |\alpha_0 - \alpha_{med}|, |\alpha_0 - \alpha_{med}|\} \tag{20}$$

so that angle differences smaller than π are always considered. The values α_0 and α_{med} are the phase of $\mathbf{V_0}$ and median phase of its neighbours $\mathbf{V_i}$. While the former is a simple arc-tangent of the components, a bit more attention must be paid for the latter. The median angle could be evaluated just taking the arc-tangent of the median of the components, but it would be biased by the magnitude of each vector. Since the magnitude fluctuation is already taken into account, neighbouring vectors need to be normalized before evaluating the angle. The median angle is then:

$$\alpha_{med} = \text{atan}\left(\frac{\text{amed}\left(\sin\theta_i, w_i\right)}{\text{amed}\left(\cos\theta_i, w_i\right)} \right) \tag{21}$$

where:

$$\theta_i = \text{atan}\left(\frac{v_i}{u_i}\right),\ \alpha_0 = \text{atan}\left(\frac{v_0}{u_0}\right) \tag{22}$$

The fluctuation of the neighbours is evaluated as before, using the median instead of the weighted median:

$$\text{fluc}^{\alpha}_{med} = \text{med}\left(\min\left\{2\pi - \left|\alpha_i - \alpha_{med}\right|, \left|\alpha_i - \alpha_{med}\right|\right\}\right) \tag{23}$$

A further discussion is required for the background error ε. As stated before, it is related to the measurement accuracy and it depends on the vector field. For PIV, this error is related to the sub-pixel fitting and is usually 0.1 pixel. This value can be used straight as a background error for the magnitude fluctuation ε^n, but it can't be used for the angle fluctuation. The minimum angle measurable depends in fact on the magnitude of the vector itself. For a vector of magnitude n_{med}, the minimum angle measurable is the angle of the triangle formed by n_{med} and $2\varepsilon^n$:

$$2\varepsilon^{\alpha} = \text{atan}\left(\frac{2\varepsilon^n}{n_{med}}\right) \tag{24}$$

Once the normalized residual has been evaluated for all the vectors of the field, non-valid vectors can be detected comparing the residuals with a threshold. The value of this threshold can be safely set to 2 [7], where higher values lead to a more stringent algorithm and vice-versa.

AVERAGED WEIGHTED MEDIAN

One of the innovations introduced in this algorithm to reduce over-detections is a new approach for the median estimation. When the normalized residual is evaluated for a scrutinised vector [7] [8], a prediction of that vector is estimated through the median of its neighbours. The more accurate the prediction, the more reliable the residual. If none of the neighbours were outliers, an interpolation of them would be the best prediction for the scrutinised vector. As already shown [1], the presence of spurious vectors among the neighbours strongly affects the estimation of the residual, therefore the median is usually proposed as most robust operator. On one hand, the median can tolerate up to 50% of the data being wrong, on the other hand the median of a sample of values is always one of the values of the sample or the average of two of them. In order to overcome this limitation, a new averaged weighted median has been introduced in order to have both a good interpolation and robustness against outliers.

The case of Figure 5 depicts a vector $\mathbf{V_0}$ and its eight closest neighbours $\mathbf{V_i}$; the figure depicts the vectors on the left and their magnitudes on the right, sorted from the smallest to the highest. Each neighbour has a different weight evaluated according to (17), normalized by the total sum of the weights and depicted through the bar width. A new parameter Δ_w called averaging interval is introduced and its range varies between 0 and 1: the new median value is the weighted average of the values falling within the interval Δ_w. Values falling on the edge of Δ_w are considered only for the part of them within Δ_w. It is clear than a value of $\Delta_w = 1$ degenerates into the weighted average, while a value of $\Delta_w = 0$ is the simple weighted median.

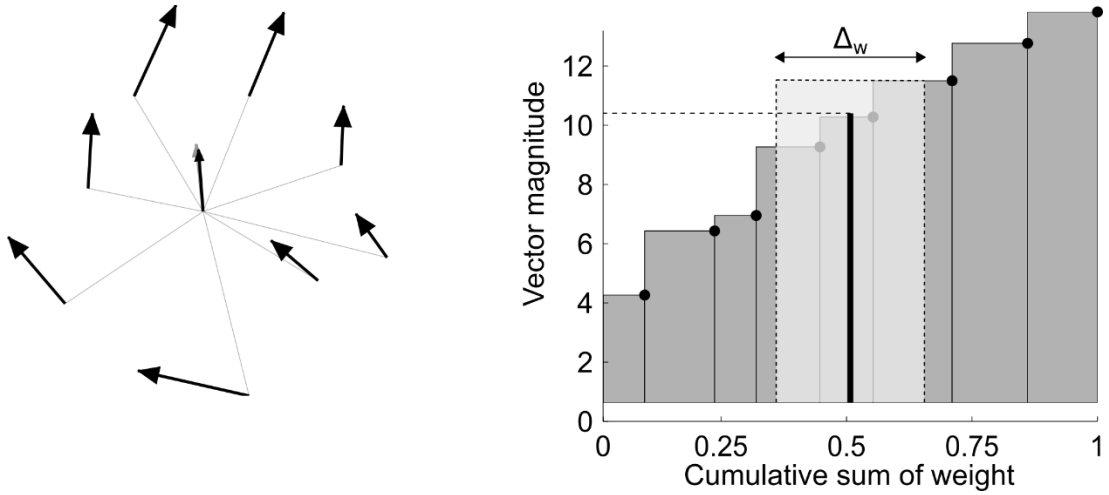


Figure 5 On the left: a vector with its eight closest neighbours and its prediction in grey. On the right: graph for evaluation of the averaged weighted median of the vectors magnitude.

The main advantage of this new median is that its value is both a weighted interpolation and robust against outliers. The vector prediction evaluated for magnitude and direction is shown in grey in Figure 5, and it is very close to the scrutinised vector.

Several tests have been performed to evaluate the effect of Δ_w on the number of under- and over-detections. Figure 6 depicts one of those tests and shows how the over-detections strongly decrease with small values of Δ_w. Concerning under-detections, there is a slight positive effect for small values of Δ_w before they start being affected negatively, therefore a value of $\Delta_w = 0.3$ has been chosen for all the tests. The experiment has been performed with a Monte Carlo simulation of 3000 independent tests, with a maximum random magnitude of 5 and clustering factor of 1, while the percentage of outliers has been kept constant at 15%.

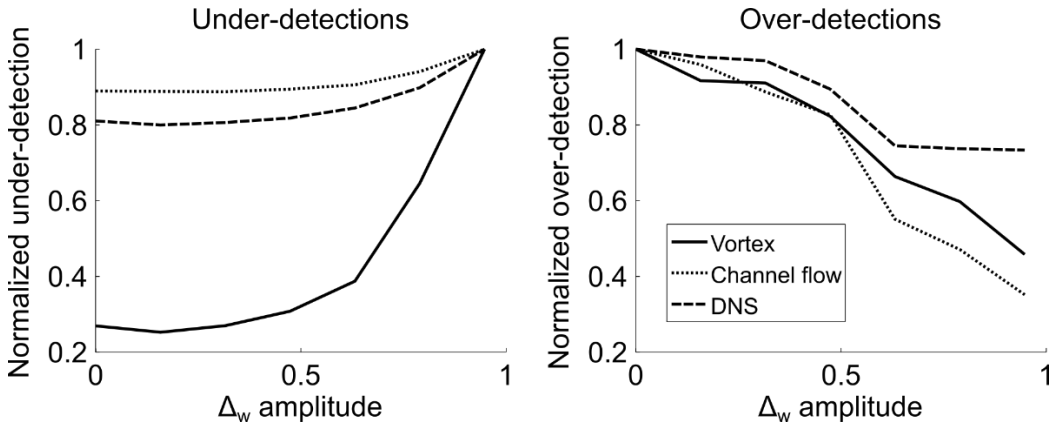


Figure 6 Result of over- and under-detections divided by the maximum value, for three test cases, changing the amplitude of the averaged weighted median.

NUMERICAL ASSESSMENT SETTINGS

In order to assess this new vector validation algorithm, correct vector fields have been corrupted with random outliers superimposed in a Monte Carlo simulation. This kind of test is quite common in vector validation algorithms [12], [5]. The outlier detection is very sensitive to the way random vectors are imposed. For our purpose, the authors argue that random outliers can be characterized by two parameters; (1) magnitude of outliers and (2) clustering factor The magnitude of outliers *M* is the maximum value of a random number multiplying the components of the correct vector being replaced. Suppose vector $\mathbf{V_0} = (u_0, v_0)$ is going to be replaced with an outlier of magnitude *M*, then $\mathbf{V_0}$ random components will be evaluated as:

$$n_0 = M\sqrt{u_0^2 + v_0^2}$$
$$u_{\text{rand}} = \frac{n_0}{\sqrt{2}}\text{Rand}(-1,1) \tag{25}$$
$$v_{\text{rand}} = \frac{n_0}{\sqrt{2}}\text{Rand}(-1,1)$$

where $\text{Rand}(-1,1)$ is a random number between -1 and 1.

The clustering factor C_f is defined as the number of neighbouring vectors randomly imposed: for instance, $C_f = 8$ means random vectors are imposed by groups of 8 neighbours each. A number of 3000 independent simulations have been accomplished for each variation of the parameters: tests has been performed varying the clustering factor with a fixed magnitude and vice-versa. For this reason, results are presented with two different types of plot: detections against clustering factor, parametrized for magnitude and detections against magnitude, parametrized for clustering factor. The authors argue that the percentage of random outliers in the field only affects the clustering factor, which becomes higher when the percentage is higher but, since the clustering factor is explicitly tested, a fixed percentage of 15% of outliers have been set for all the tests.

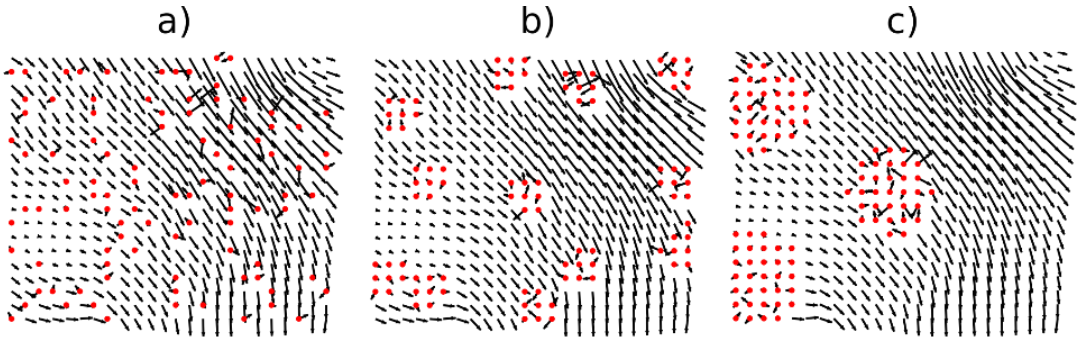


Figure 7 Part of a sample vector field with random outliers over-imposed by clustering factor of: 1 (a), 8 (b) and 32 (c). The magnitude M is set to 1.

In order to quantify the results of each validation test, erroneous detection are shown as number of over-detections N_w (correct vectors detected as outliers) and the number of under-detections N_m (random outliers detected as correct vectors). Number of detections are conveniently normalized by the total number of vectors (valid and invalid) and the imposed number of spurious vectors in order to get the ratio of over-detection , and the ratio of under-detection as it follows:

$$R_o = \frac{N_w}{N_t} \text{ and } R_u = \frac{N_m}{N_i} = \frac{N_m}{0.15 \cdot N_t} \tag{26}$$

In addition to this, the ratio between N_m and N_t, denoted by R_u^*, allows a comparison between the over- and under-detections.

The aim of the vector validation should be to minimize both the over- and under-detections: algorithms producing zero over-detection but 100% over-detection and vice-versa are obviously useless. Results of the new algorithm are compared with the Normalized Median Test (NMT) [7] and the Distance Weighted Normalized Median Test (DW-NMT) [8]. In order to stress the importance of the various innovations introduced in this algorithm, a modified version of NMT and DW-NMT with a variable number of neighbours is proposed as well, and referred as Adaptive Normalized Median Test (ANMT) and Adaptive Distance Weighted Normalized Median Test (ADW-NMT). The new method presented in this paper is referred as Adaptive Weighted Angle and Magnitude Threshold (AWAMT). Values of over- and under-detections are shown as percentage, the former normalized by total number of vectors in the field, the latter by the number of random vector imposed. Under-detection are also presented with the same scale of over-detection to allow a comparison.

Regarding the origin of the vector fields, depicted in Figure 8, two of the three test cases are the result of an adaptive

PIV algorithm [13] for synthetic images of a time-resolved channel flow[2] and an experimental vortex[3] from the PIV challenges. These vector fields have been analysed with specially tuned settings, without any vector validation algorithm, and have been manually inspected for outliers. The third test case is a flow field resulting from DNS velocity data[4], obtained from the Johns Hopkins Turbulence Databases [14]. The first two velocity fields lie on a scattered grid, while the latter lies on a structured grid.

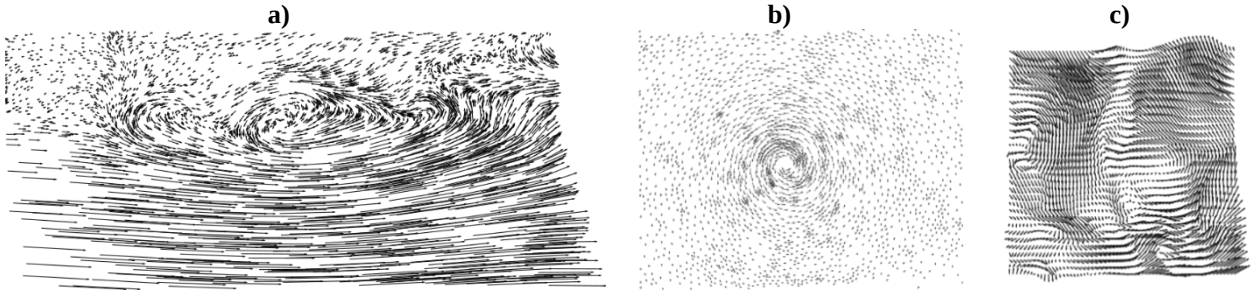


Figure 8 Flow test cases used to assess the vector validation algorithm: a) PIV unstructured channel flow b) PIV unstructured vortex and c) DNS turbulent flow

[2] PIV Challenge (Sept.19-20, 2005, Pasadena, USA), CASE B Time-resolved Channel Flow (Synthetic image, B_037/B_038)
[3] PIV Challenge (Sept.14-15, 2001, Göttingen, Germany), CASE A Strong Vortex (experimental image provided by Kaehler)
[4] Dataset: *isotropic1024coarse*; x-range: (1.1781, 1.5650); y-range: (1.1781, 1.5650); z-offset: 0.0055; time: 1.9013

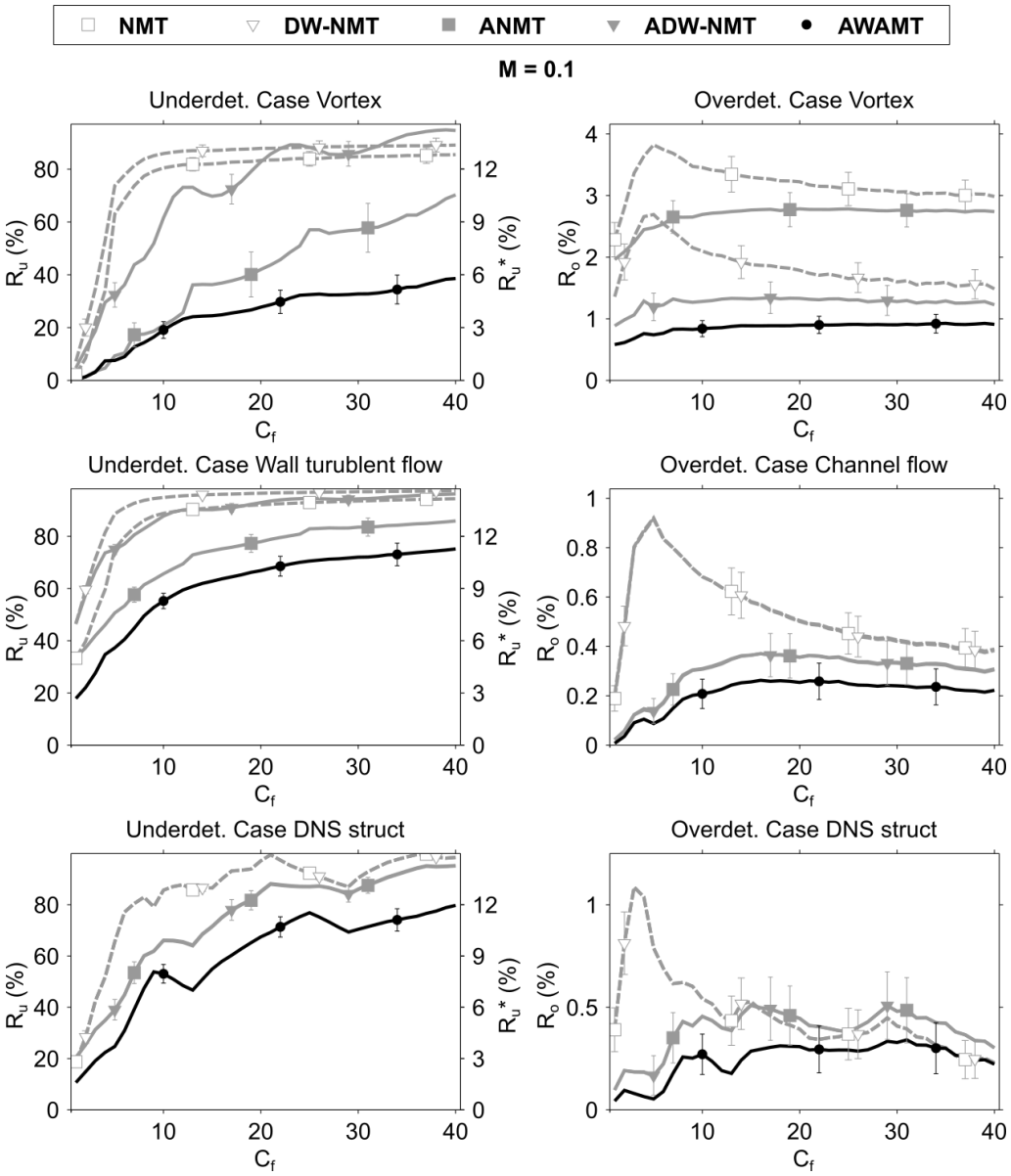


Figure 9 Result of simulation varying the cluster size at a fixed magnitude of 0.1

DISCUSSION OF RESULTS

Figure 9, 10 and 11 show the quantity of under- and over-detection in a case where the magnitude of random vectors has been fixed (respectively to 0.1, 1 and 10), while the clustering factor has been changed. The dashed lines depict NMT and DW-NMT methods, the grey solid lines depicts their respective version with a variable number of neighbours implemented, while the black solid line depict the new method AWAMT. The first result that can be noticed in all of the figures is that the variable number of neighbours allows a great enhancement of both under- and over-detections in all of the test cases, even without any further improvement.

For small random outliers in Figure 9, all the results show that the new AWAMT method is able to detect outliers better than other methods reaching a lower value both for over- and under-detections. In this first case the difference between NMT and DW-NMT methods and their variable neighbours implementation is remarkable yet not very strong: this is due to the very small magnitude of outliers that are shrunk more than 10 times which does not allow a reliable estimation of the vector coherency. On the other hand, the magnitude-orientation residual elaborated in the new AWAMT algorithm allows a better detection of outliers even though they are very small.

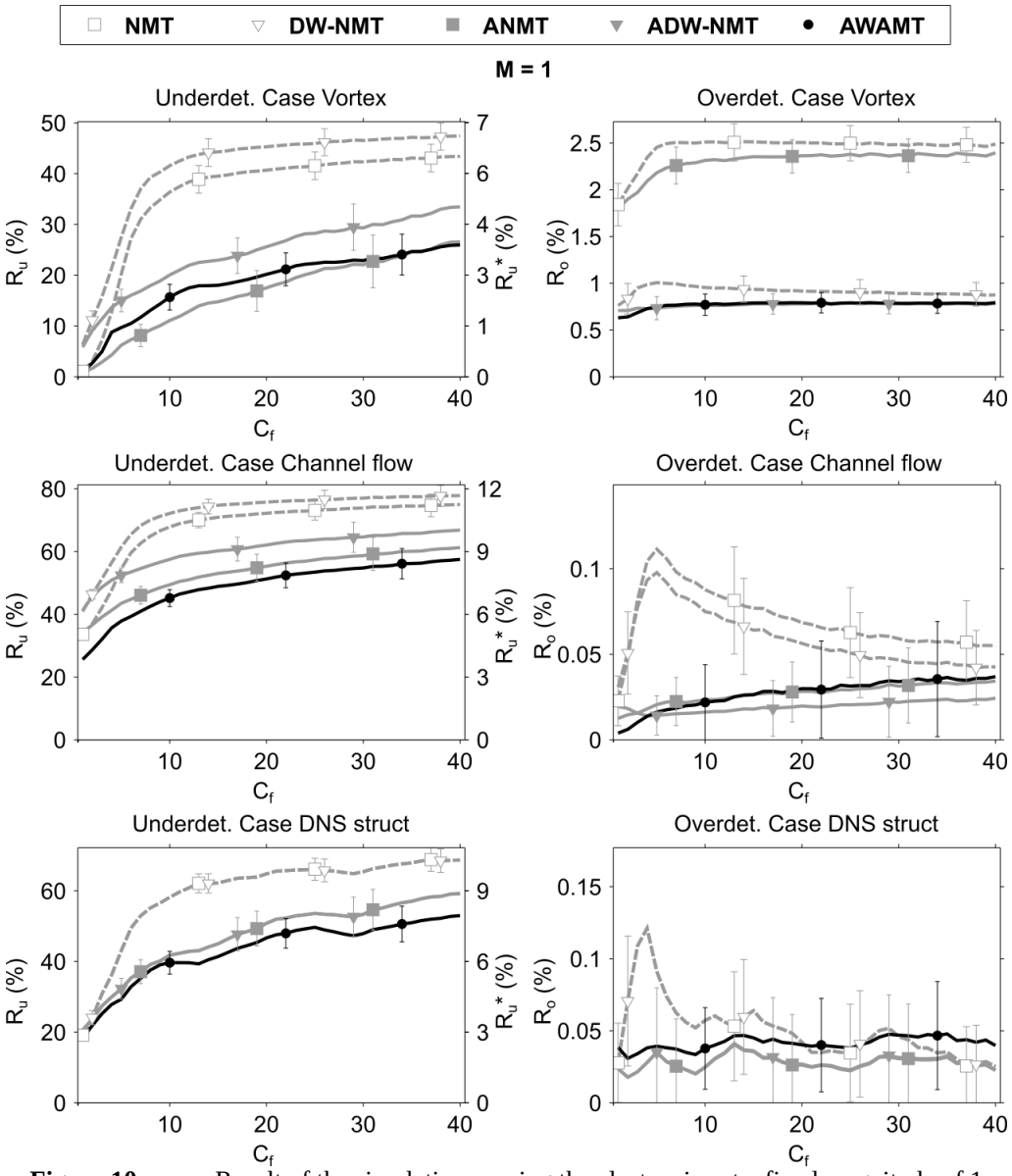


Figure 10 Result of the simulation varying the cluster size at a fixed magnitude of 1

Figure 10 depicts a case where outliers have magnitude 1, meaning their maximum magnitude is the same of the correct vectors being replaced. Also in this case, the new method shows overall a better behaviour: in the case of the vortex ANMT have slightly less under-detections, yet much more over-detections, while ADW-NMT has slightly the same over-detections, yet more under-detections. From this magnitude on, the difference between the NMT and DW-NMT standard algorithms and their variable neighbour implementations becomes even more remarkable, showing the better influence of this part of the algorithm on the detection.

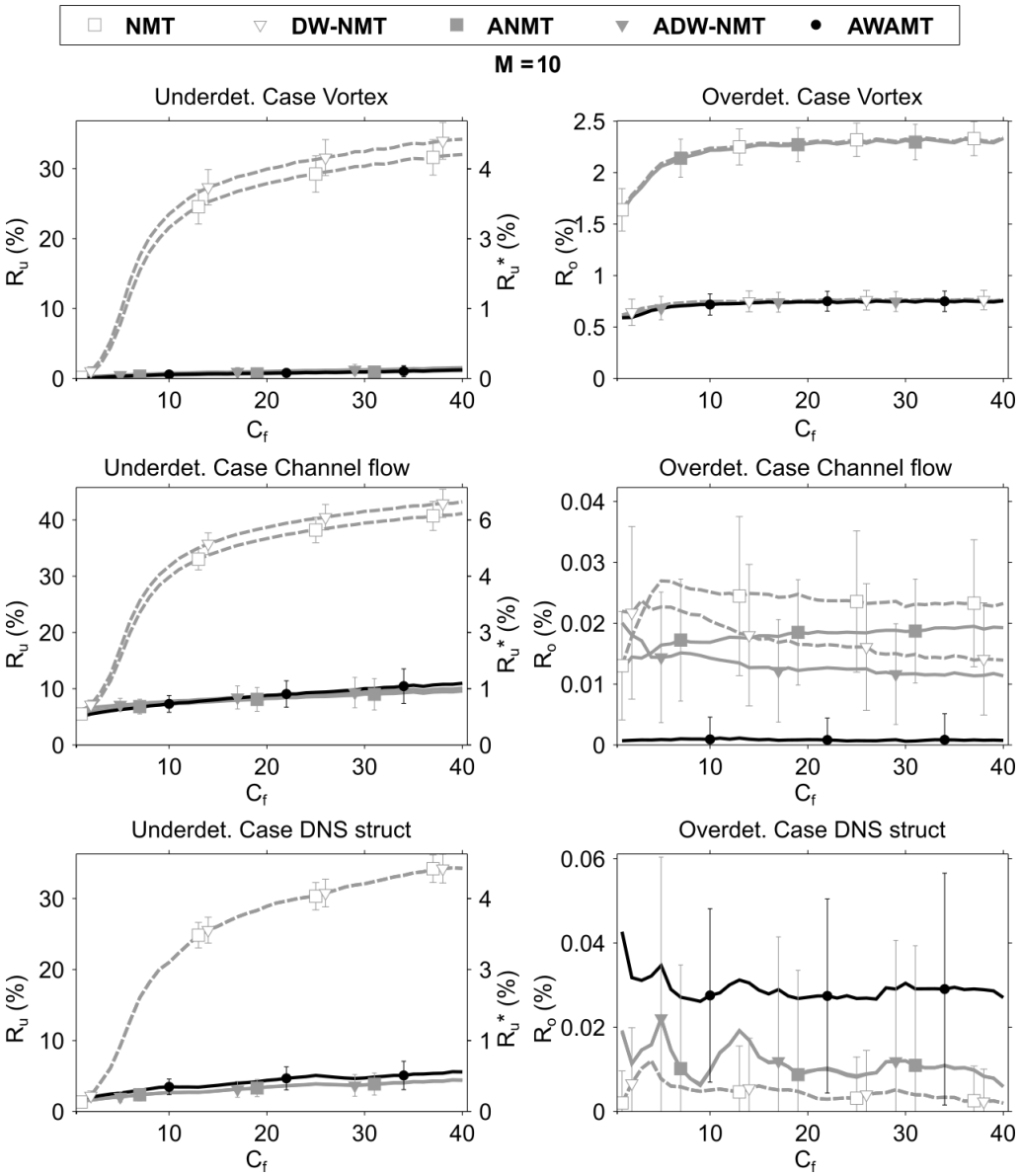


Figure 11 Result of the simulation varying the cluster size at a fixed magnitude of 10

The last case in Figure 11 depicts the results where outliers have a maximum magnitude that is ten times bigger than the actual vector field. As already stated, in this case the difference between the variable-neighbours algorithms and the constant-neighbours algorithm becomes even more remarkable. This behaviour is corroborated by results shown in Figure 12, where the clustering factor is fixed to 8 and the magnitude of outliers is changed continuously: when the outlier intensity becomes very high all the variable-neighbours algorithms tend asymptotically to a constant value both for over- and under-detections that is much smaller than the constant-neighbours algorithm.

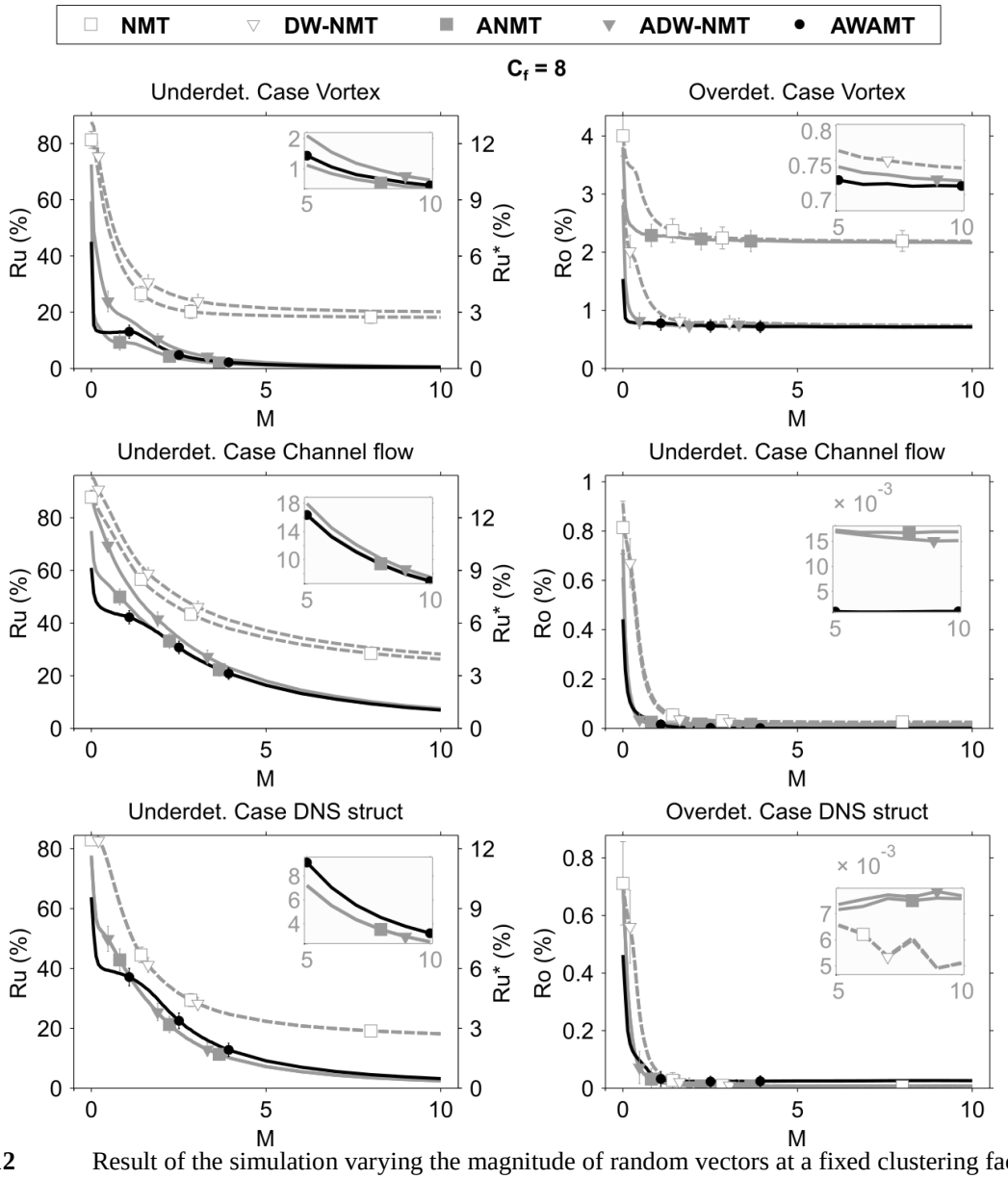


Figure 12 Result of the simulation varying the magnitude of random vectors at a fixed clustering factor of 8

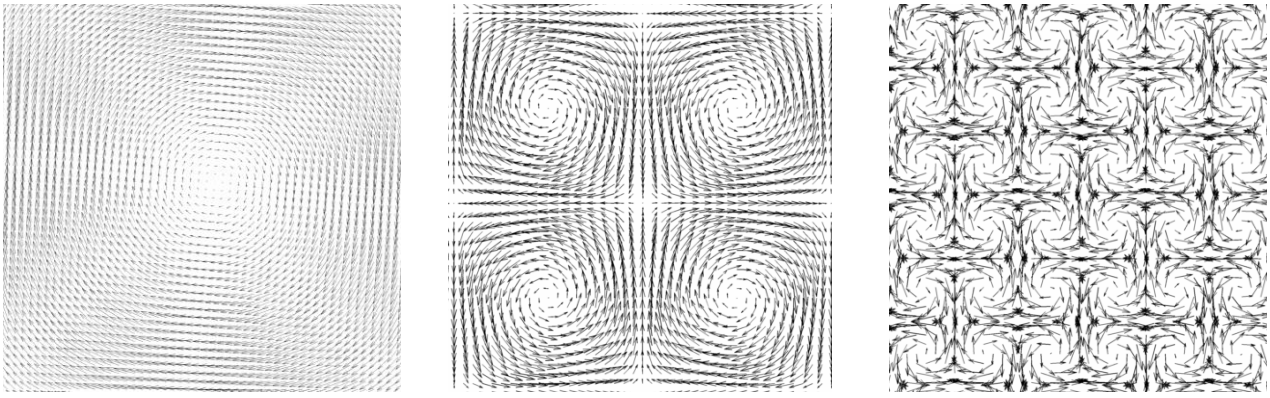
The last case in Figure 12 shows a slightly higher number of both over- and under-detections for the new proposed algorithm AWAMT that, although small, must be discussed. This result is not related to the use of a structured grid in the test case but with the flow itself. Due to the very small scale of the original flow, it has been sub-sampled selecting every other vector, in order to get a velocity field large enough to contain some vortices yet having a reasonable number of total vectors suitable for the Monte Carlo simulation. This process has led to a vector field containing weird vectors that, indeed, seem to be incoherent. The algorithm detects those vectors as outliers, but they result as over-detected because they have not been imposed as outliers. Further tests evaluated with classic PIV data lying on a structured grid corroborates the results of the previous test cases.

SYNTHETIC OVER-DETECTION TEST

To further stress the improvement of AWAMT in terms of over-detection, an additional test on a synthetic flow field is proposed in this section. The test has been performed on an outlier-free velocity field described by:

$$u = -A\sin\left(\frac{2\pi}{\lambda}y\right)\cdot\cos\left(\frac{2\pi}{\lambda}x\right)$$
$$v = A\sin\left(\frac{2\pi}{\lambda}x\right)\cdot\cos\left(\frac{2\pi}{\lambda}y\right) \quad (27)$$

The flow described by (27) is depicted in Figure 13, where λ indicates twice the distance between two consecutive vortices. No outliers were imposed to the data: the authentic velocity field has been analysed with three different validation techniques to solely study the over-detection of the outlier detection methods discussed above.



(a) **(b)** **(c)**

Figure 13 Array of vortices used for the synthetic over-detection test in case of **(a)** $WS/\lambda = 0.015$ **(b)** $WS/\lambda = 0.05$ **(c)** $WS/\lambda = 0.15$

The analysis has been performed counting the number of vectors detected: since no outliers were imposed, all the vectors detected as wrong are considered contribute to over-detection. In order to avoid the effect of vortex sub-sampling, multiple analyses for multiple random shifts of the array were performed and the average results are presented in Figure 14. The plot clearly shows how AWAMT is less affected by the problem of over-detection, compared with NMT and DW-NMT As the size of the vortices reduces, stronger gradients in the velocity field appear and AWAMT is capable of reducing the overdetection ratio from 11% to 7%. As gradients diminish in strength, R_o generally decreases although AWAMT continues to perform better. These findings corroborate the results obtained in the previous section. It is also interesting to note that DW-NMT has a higher over-detection ratio at small WS/λ compared to NMT.

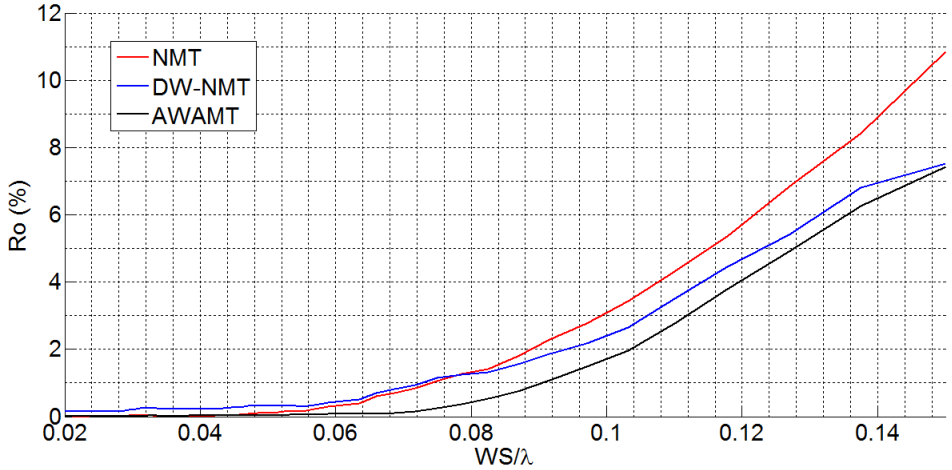
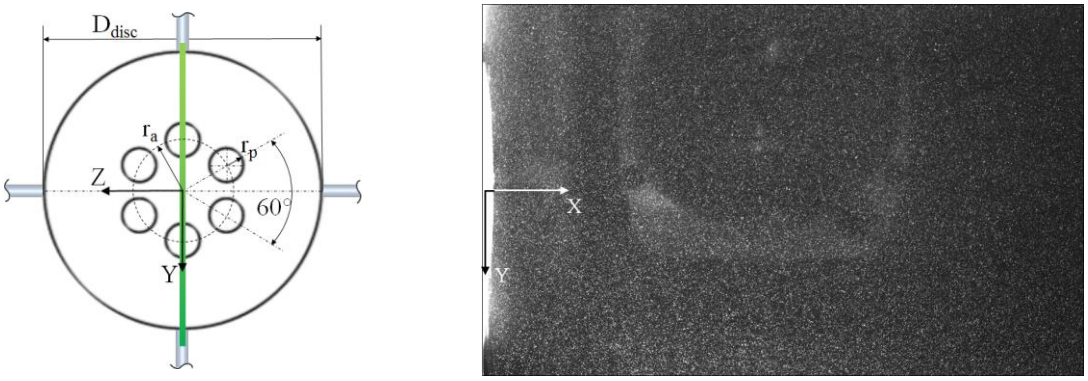


Figure 14 Ratio of over-detection against vortex size, for an outlier-free vortex array displacement field, analysed with NMT (—), DW-NMT (—) and AWAMT (—)

APPLICATION TO PIV

Algorithms for the vector validation can have a strong influence on the outcome of a PIV analysis, especially because of its iterative implementation. To assess the new vector validation algorithm in a real application it has been implemented in a PIV routine and used to analyse experimental particle images. Measurements were conducted behind a circular disc placed perpendicular to a 30 m/s flow in the low turbulence wind tunnel of the University of Bristol. This tunnel attains turbulence levels below 0.05% and has an octagonal test section of 0.8 m×0.6 m. The disc had a thickness of 6mm and a diameter of 6cm resulting in a negligible blockage of 0.16% at a diameter-based Reynolds number ReD of 118·103. Six perforations, each with a radius of 3.87mm, were placed at a radius of 1.08cm to establish a porosity β (=open/closed area) of 0.11.

PIV measurements were performed with a two-dimensional two-component Dantec Dynamics PIV system. Seeding was generated by atomizing a mixture of PEG-80 and water producing 1μm tracer particles. Illumination was provided by a Litron 200mJ laser, which was optically transformed into a 1mm thick laser sheet across two pores in the symmetry plane of the disc.



(a) **(b)**

Figure 15 (a) Disc used in the experiment for the PIV analysis. (b) Exemplary PIV image illustrating the reflections off the disc surface

In order to ensure a fair comparison, vectors detected as outliers are always replaced with the same technique, substituting the wrong components with the median of their closest neighbours, so that the only difference between the two implementations is the number of vectors detected. A set of 430 image pairs were analysed with both NMT and AWAMT validation algorithm and a quiver plot for the average of the velocity field is shown in Figure 16.

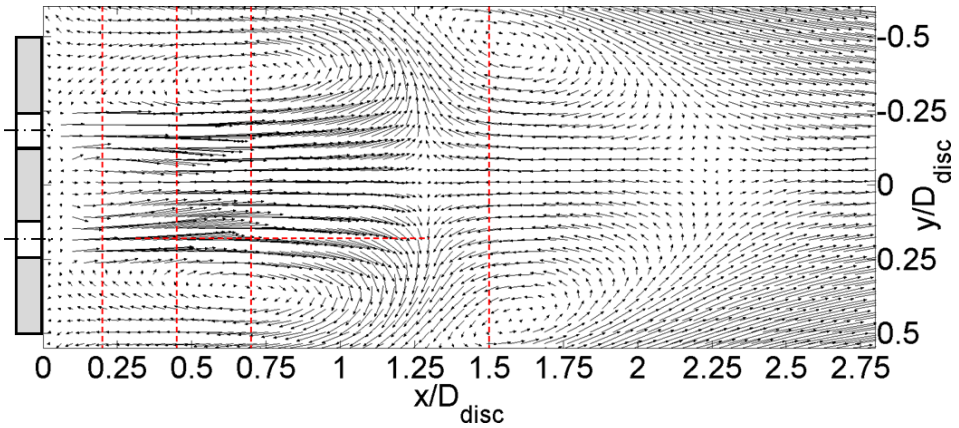


Figure 16 Quiver plot of the resulting velocity field from the analysis of 430 synthetic images. The disc is placed on the left side of the image and the flow is going from left to right

The number of replacements for each vector, in the final iteration of the PIV process, has been stored for both the algorithms, and the result for all the images is shown in Figure 17. Images clearly show that the new AWAMT has a much lower number of detections, especially in the areas were gradients are stronger. It can also be noted that NMT systematically over-detects vectors on the edges of the image: on the borders, in fact, less information on velocity is available and NMT tends to over-detect vectors. In the new AWAMT this issue is mitigated by the variable number of neighbours and the better estimation of a prediction through the new averaged weighted median.

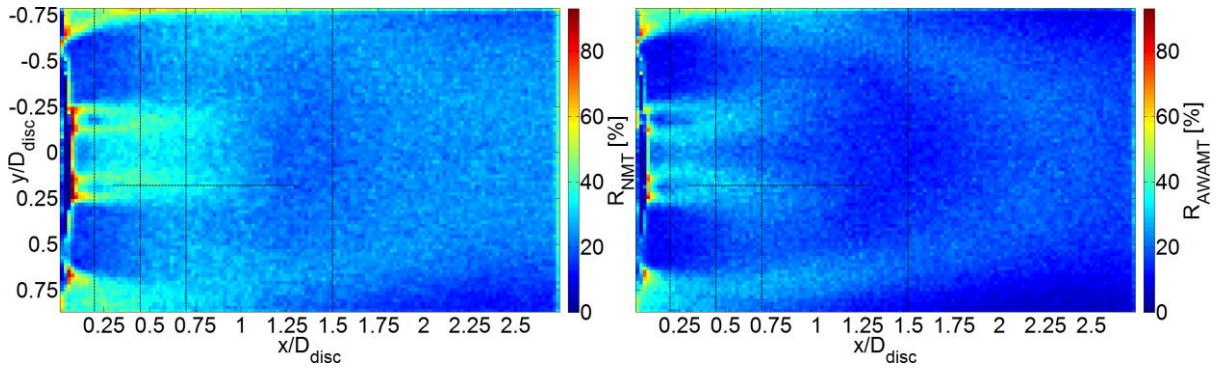
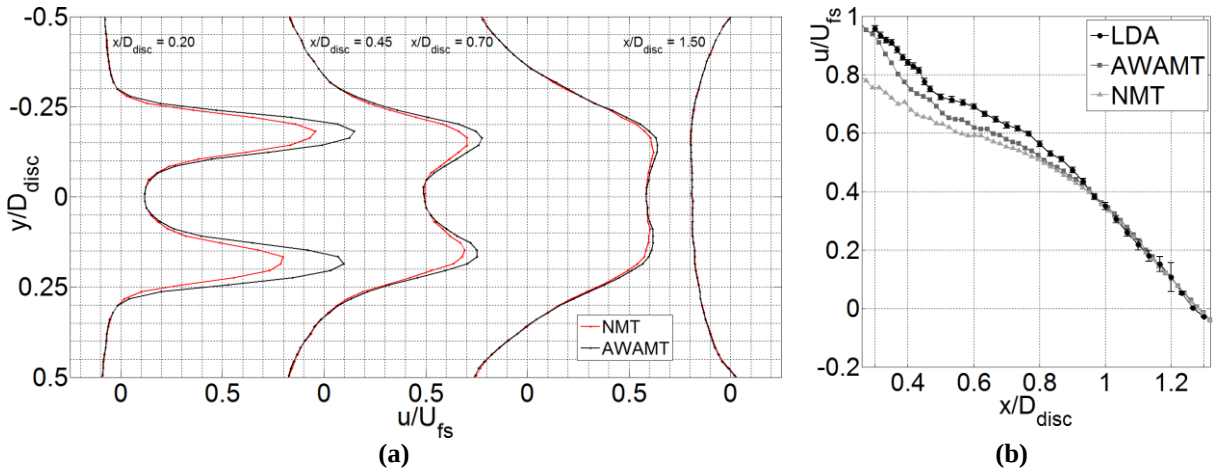


Figure 17 Sum of the number of replacements for each vector of the field across 430 synthetic images analysed

The advantage of AWAMT in terms of measurement accuracy is highlighted in Figure 18, where ensemble averaged profiles in horizontal velocity are juxtaposed. While velocity profiles at x/D_{disc}=0.2 are on a par in the regions $|y/D_{disc}|$<0.1 and $|y/D_{disc}|$>0.25, a larger disparity between the two validation methodologies can be observed in jet centreline velocities. Whereas AWAMT yields a peak velocity ratio of approximately 1.1, NMT predicts a ratio of approximately 0.8. The observed modulation in jet velocity thus shows a strong correlation with R as the inherent vector re-interpolation will introduce a smoothening effect. Travelling downstream, the two pore jets increase in width and the centrelines arc inwards to merge at x/D_{disc}=0.7 similar to well-documented parallel jet theory [15]. Velocity gradients gradually reduce in strength and the discrepancy in jet centreline velocity between the validation routines reduces from 0.64 for AWAMT compared to 0.61 with NMT at x/D_{disc}≈0.7.



(a) **(b)**

Figure 18 (a) Profiles of the horizontal velocity component extracted along the lines indicated in Figure 16

To provide an estimate of the underlying true flow field, PIV data extracted along y/D_{disc}=0.18 is superimposed with results obtained from a two-component Dantec Dynamics Laser Doppler Anemometry (LDA) system operating in crossed beam mode (Figure 18b). The LDA measurement volume extended approximately 0.17mm (~$0.0028 \cdot D_{disc}$) in streamwise normal direction providing spatially highly resolved measurement data. Because of the inherent beam alignment measurements closest to the disc were restricted to 0.3 disc diameters. Velocity samples were spaced approximately $0.017 \cdot D_{disc}$ up to 1.3 diameters downstream. At each measurement location velocity statistics were evaluated on the basis of typically 6000 instantaneous samples sampled at 4kHz.
The observable tendency in LDA data confirms velocity exiting from the pores can surpass the freestream condition, suggesting the holes to act as contractions. Towards the disc the magnitude of the average LDA horizontal velocity component is seen to exceed the PIV data which is caused by the bias of the PIV data towards lower velocities.

CONCLUSIONS

Velocity fields obtained from PIV image analysis techniques are always contaminated with erroneous vectors and such outliers often appear in clusters as a result of underlying degraded image quality or strong gradients in flow velocity. Existing validation methodologies for instantaneous PIV velocity fields are commonly based on comparison of the scrutinised vector with its immediate neighbourhood. As a result such methods are unable to detect false vectors when clustered and are moreover prone to mistakenly invalidate correct vectors. For this reason a novel adaptive method for outlier detection has been proposed in this paper with the aim to render validation processes more robust in the presence

of outlier clusters. The detection of false vectors will thereby be improved and over-detection can be reduced without the need to fine-tune inherent parameters.
The proposed method emulates the process of outlier detection in human vision, whereby the considered neighbourhood for comparison is a-priori extended until the database is sufficiently reliable for a posteriori validation tasks. Selection of the appropriate vicinity is dictated by a measure of coherency. The latter is quantified as the discrepancy between local velocity values and a parabolic regression. For each vector the neighbourhood is automatically enlarged until at least half the enclosed vectors are coherent. To further improve the validation algorithm, vector comparison is performed on the basis of magnitude and direction instead of the traditional horizontal and vertical vector components. To limit the potential diversity in vector direction, the acceptable background fluctuation level is automatically adjusted to the vector magnitude and constitutes a second feature of adaptivity. Moreover, applicability to both structured and unstructured data grids is ensured by the implementation of a distance-based Gaussian weighting system.
The algorithm has been assessed with Monte-Carlo simulations using three flow fields; an isolated vortex, a channel flow and a DNS simulation of isotropic turbulence. The former two cases were obtained from the PIV challenges while the latter originated from the John Hopkins Turbulence Database. The flow fields were contaminated with outliers of varying magnitude and degree of clustering. The common outlier detection schemes resulted in high numbers of undetected outliers (under-detection) and number of wrongly invalidated correct vectors (over-detection). Depending on the amount of clustering and outlier magnitude as much as 80% of the spurious vectors could remain undetected, while 3% of the total vectors could be over-detected. Implementation of the coherency adaptivity dramatically improved the outlier detection, potentially reducing the under-detection by as much as one-fourth for small outlier magnitudes or even one-tenth for larger magnitudes. These findings advocate coherency adaptivity to be a powerful tool to improve the performance of existing validation routines even in the presence of outlier clusters. The concept is computationally simple and can be easily implemented. Validation on the basis of angle and magnitude enabled a further lowering of the missed outliers and mistaken outliers though improvements were only observed at low outlier magnitudes. Overall the proposed validation method was the most robust and general without any reliance on user-defined parameters. Validation performances consistently surpassed the traditional routines and were better or at least on a par with the adaptivity enhanced methodologies.
When implemented in a standard PIV image analysis process and applied to experimental PIV images of a porous disc's near-wake, the novel outlier detection routine was shown to drastically reduce the amount of over-detection. As a consequence, a gain in spatial resolution in term of velocity gradient across the shear layer was noticeable as well as an increased flow accuracy.

ACKNOWLEDGEMENT

This work has been funded by the UK's Engineering and Physical Sciences Research Council (EPSRC-EP/L010755/1). Their support is greatly appreciated.

REFERENCES

[1] J. Westerweel "Efficient detection of spurious vectors in particle image velocimetry data" Exp. Fluids vol. 16 (1994) pp. 263–247

[2] R. D. Keane, R. J. Adrian "Optimization of particle image velocimeters. Part I : Double pulsed systems" Meas. Sci. Technol. vol. 1 (1990) pp. 1202–1215

[3] J. H. Sun, D. a. Yates, D. E. Winterbone "Measurement of the flow field in a diesel engine combustion chamber after combustion by cross-correlation of high-speed photographs" Exp. Fluids vol. 20, no. 5 (1996) pp. 335–345

[4] X. Song, F. Yamamoto, M. Iguchi, Y. Murai "A new tracking algorithm of PIV and removal of spurious vectors using Delaunay tessellation" Exp. Fluids vol. 26, no. 4 (1999) pp. 371–380

[5] D. F. Liang, C. B. Jiang, Y. L. Li "Cellular neural network to detect spurious vectors in PIV data" Exp. Fluids vol. 34, no. 1 (2003) pp. 52–62

[6] B. Reiz, S. Pongor "Psychologically inspired, rule-based outlier detection in noisy data" Proc. - 13th Int. Symp. Symb. Numer. Algorithms Sci. Comput. SYNASC 2011 (2011) pp. 131–136

[7] J. Westerweel, F. Scarano "Universal outlier detection for PIV data" Exp. Fluids vol. 39, no. 6 (Aug. 2005) pp. 1096–1100

[8] J. Duncan, D. Dabiri, J. Hove, M. Gharib "Universal outlier detection for particle image velocimetry (PIV) and particle tracking velocimetry (PTV) data" Meas. Sci. Technol. vol. 21, no. 5 (May 2010)

[9] R. Theunissen, F. Scarano, M. L. Riethmuller “Spatially adaptive PIV interrogation based on data ensemble” Exp. Fluids vol. 48 (2010) pp. 875–887

[10] J. Nogueira, a Lecuona, P. a Rodríguez “Data validation, false vectors correction and derived magnitudes calculation on PIV data” Meas. Sci. Technol. vol. 8, no. 12 (1999) pp. 1493–1501

[11] J. C. Agüí, J. Jiménez “On the performance of particle tracking” J. Fluid Mech. vol. 185 (1987) pp. 447–468

[12] A. M. Shinneeb, J. D. Bugg, R. Balachandar “Variable threshold outlier identification in PIV data” Meas. Sci. Technol. vol. 15, no. 9 (2004) pp. 1722–1732

[13] R. Theunissen, F. Scarano, M. L. Riethmuller “On improvement of PIV image interrogation near stationary interfaces” Exp. Fluids vol. 45 (2008) pp. 557–572

[14] Y. Li, E. Perlman, M. Wan, Y. Yang, C. Meneveau, R. Burns, S. Chen, A. Szalay, G. Eyink “A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence” J. Turbul. vol. 9, no. 31 (2008) pp. 1–29

[15] E. a. Anderson, R. E. Spall “Experimental and Numerical Investigation of Two-Dimensional Parallel Jets” J. Fluids Eng. vol. 123, no. 2 (2001) p. 401